

RIDGE REGRESSION (L2 NORM REGULARISED REGRESSION)

* WHY? PENALTY ON WEIGHTS (θ_i^2) as LARGE

$\theta_i \rightarrow$ OVERTFITTING.

* OBJECTIVE

MINIMIZE

$$(y - X\theta)^T (y - X\theta)$$

Objective of Least Squares

s.t.

$$\theta^T \theta \leq S$$

Constrain θ_i^2

EQUIVALENT TO:

$$\text{MINIMIZE } (y - X\theta)^T (y - X\theta) + \lambda \theta^T \theta$$

HOW? KKT CONDITIONS

$$\text{MINIMIZE } (y - X\theta)^T (y - X\theta)$$

s.t.

$$\theta^T \theta - s \leq 0$$

$$L(\theta, \mu) = (y - X\theta)^T (y - X\theta) + \mu (\theta^T \theta - s)$$

$\mu \geq 0$

COMPLEMENTARY SLACKNESS

$$\mu (\theta^T \theta - s) = 0$$

if $\mu = 0$

NO
REGULARIZATION

~~X~~

NO EFFECT OF CONSTRAINT

$$\mu \neq 0$$

$$\Rightarrow \theta^T \theta - s = 0$$

RIDGE

OBJECTIVE

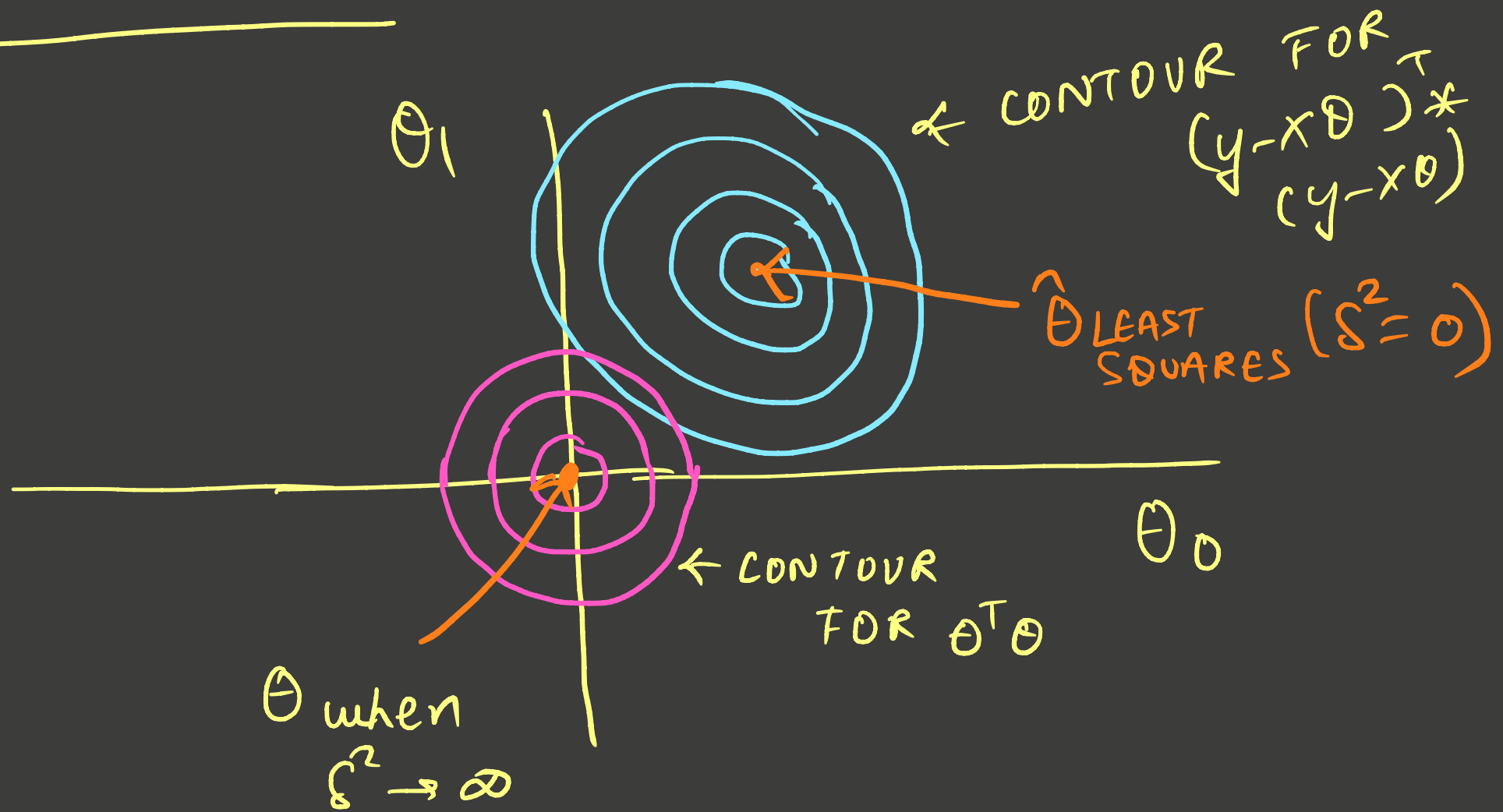
$$\min_{\theta} (y - X\theta)^T (y - X\theta) + \delta^2 \theta^T \theta$$

$$\frac{\partial L(\theta, \mu)}{\partial \theta} = 0 \Rightarrow \underbrace{\left[\frac{\partial}{\partial \theta} \left\{ y^T y - 2y^T X\theta + \theta^T X^T X\theta \right\}}_{\text{or}} \right]}_{\text{or}} + \frac{\partial}{\partial \theta} \delta^2 \theta^T \theta = 0$$

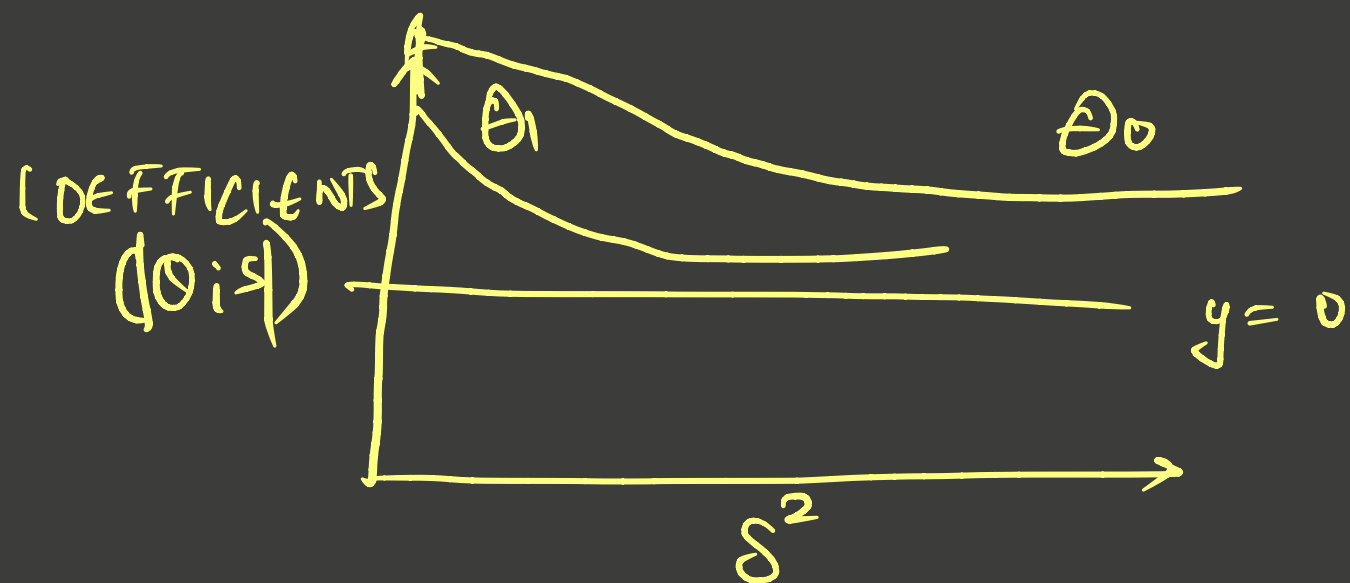
$$\Rightarrow -X^T y + (X^T X + \delta^2 I) \theta = 0$$

$$\Rightarrow \theta^* = \underbrace{(X^T X + \delta^2 I)^{-1}}_{\text{DIFFERENCE}} X^T y$$

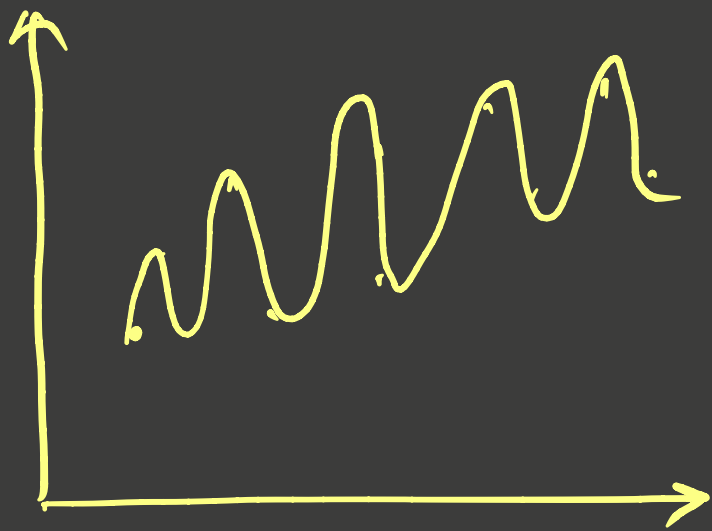
GEOMETRIC MEANING



REGULARISATION PATH



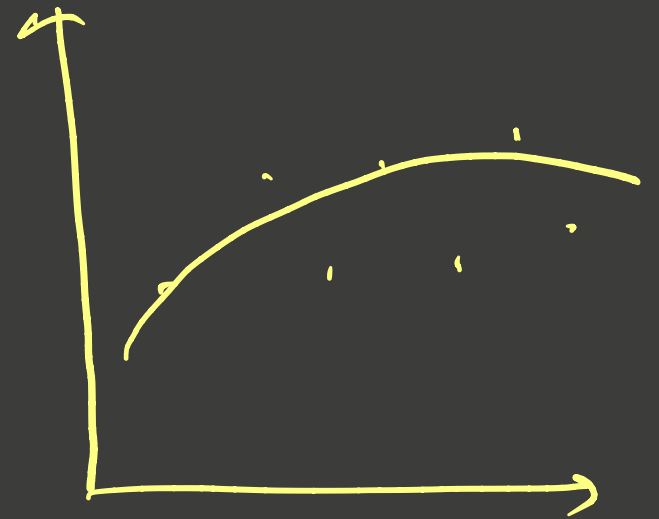
BIAS/VARIANCE W.R.T. δ^2 (REGULARISATION)



FIT HIGH ORDER
POLYNOMIAL \Rightarrow
HIGH VARIANCE

$$\delta^2 \rightarrow 0$$

\Rightarrow OBJECTIVE
ONLY FIT
DATA



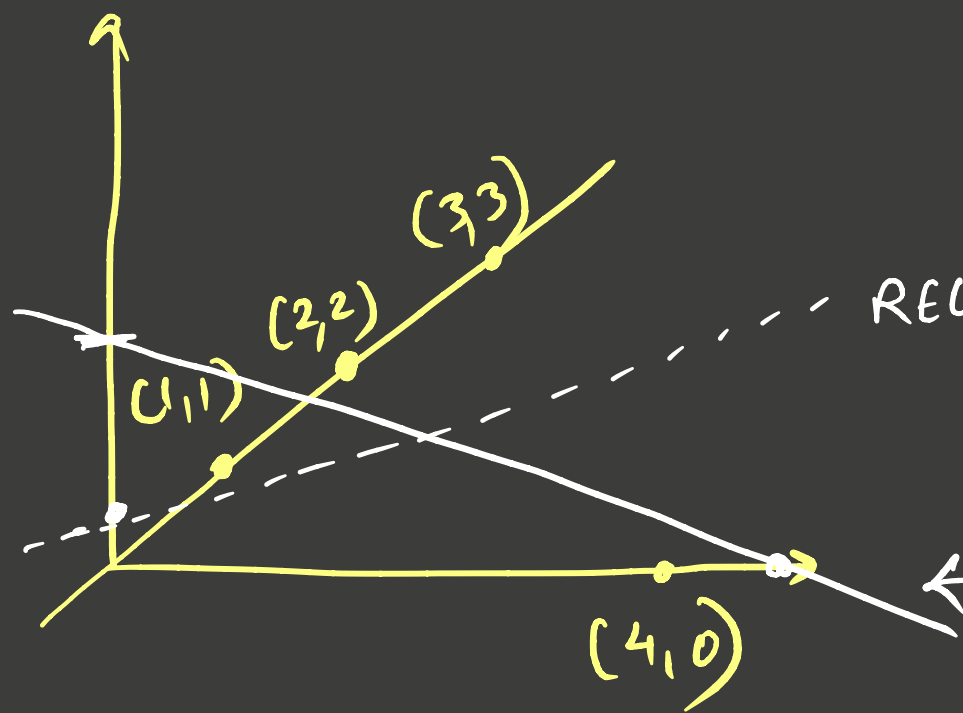
HIGH
BIAS
(θ 's SHOULD
BE SMALL)

$$\delta^2 \rightarrow \text{HIGH}$$

OBJECTIVE:

BOTH

- ① FIT DATA
- ② CONSTRAINT θ_i^2



Q.) SOLVE REGULARISED
& ($\delta^2 = 2$
RIDGE)
UNREGULARISED
REGRESSION

← UNREGULARISED
FIT

REGULARISED

$$\hat{\theta} = (X^T X + \delta^2 I)^{-1} (X^T y)$$

$$= \left\{ \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \right\} \begin{bmatrix} 6 \\ 14 \end{bmatrix}$$

$$\hat{\theta} \approx \begin{bmatrix} 0.5 \\ 0.4 \end{bmatrix}$$

RIDGE FIT

$$\hat{y} = 0.5 + 0.4x$$

$$\theta^T \theta = 0.25 + 0.16 \approx 0.4$$

UNREGULARISED CASE

UNREG. FIT $\theta^T \theta \approx 4$

$$\hat{y} = 2 - \frac{1}{5}x$$

MODIFY FORMULAT^N FOR RIDGE

S.T.

NO PENALTY ON θ_0

$$\hat{\theta} = (X^T X + \lambda^2 I')^{-1} X^T y$$

where

$$I' = \begin{bmatrix} 0 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}$$